Integrating Geographical Information Systems and Grid Applications

Marlon Pierce (mpierce@cs.indiana.edu)
Contributions: Ahmet Sayar, Galip Aydin, Mehmet Aktas, Harshawardhan Gadgil, Zhigang Qi
Community Grids Lab
Indiana University
Project Funding: NASA AIST

www.servogrid.org/slide/iSERVO/ESTO2006

QuakeSim/SERVO Grid Architectural Highlights

- Use of Grid of Grids architecture
  - Based on web services
  - “data grid” services combined with “execution grid” services developed in earlier CT project.
  - Information and orchestration services integrate data sources with applications
- Data grid services based on Geographical Information System standards
  - Web Feature and Web Map Services.
  - Architecturally interesting use of “capabilities” metadata that we are exploring as both a federating and workflow planning technology
- Use of messaging infrastructure (NaradaBrokering) for real time Grids

Key Problems

- Many of the SERVO application take geospatial and geo-temporal data as input
  - Fault models, GPS stations, seismic events.
- We need to provide programmatic access to these remote data sources.
  - Ideally, we don't provide all of the services.
- Geographically distributed applications, data sources, and other tools must be combined in a lightweight, open fashion
  - Workflow, information services
- Can we do this in real time?
- This is a Grid problem.

SERVO/QuakeSim Services Eye Chart

Service	Description
Job Management	SERVO wraps Apache Ant as a web service and uses it to launch jobs. For a particular application, we design a build.xml template. The interface is simply a string array of build properties called for by the template. We've also built a simple generic "template engine" version of this.
Specific Applications: Virtual California, Geofest, Park, RDAHMM ..	These can be all launched by a single Job Management service or by custom instances of this with metadata preset to a particular application
Context Data Service	We store information gathered from users' interactions with the portal interface in a generic, recursively defined XML data structure. Typically we store input parameters and choices made by the user so that we can recover and reload these later. We also use this for monitoring remote workflows. We have devoted considerable effort into developing WS-Context to support the generalization of this initial simple service.
Application and Host Metadata Service	We have an Application and a Host Descriptor service based on XML schema descriptors. Portlet interfaces allow code administrators to make applications available through the browser.
File Services	We built a file web service that could do uploads, downloads, and crossloads between different services. Clearly this supports specific operations such as file browsing, creation, deletion and copying.
Portal	We use an OGCE based portal based on portlet architecture
Authentication and Authorization	This uses capabilities built into portal. Note that simulations are typically performed on machines where user has accounts while data services are shared for read access
Information Service	We have built data model extensions to UDDI to support XPath queries over Geographical Information System capability.xml files. This is designed to replace OGC (Open Geospatial Consortium) Web registry service
Web Map Service	We built a Web Service version of this Open Geospatial Consortium specification. The WMS constructs images out of abstract feature descriptions.
Web Feature Service	We've built a Web Service version of this OGC standard. We've extended it to support data streaming for increased performance.

Service Eye Chart Continued

Workflow/Monitoring/Management Services	The HPSearch project uses HPSearch Web Services to execute JavaScript workflow descriptions. It has more recently been revised to support WS-Management and to support both workflow (where there are many alternatives) and system management (where there is less work). Management functions include life cycle of services and QoS for inter-service links
Sensor Grid Services	We are developing infrastructure to support streaming GPS signals and their successive filtering into different formats. This is built over NaradaBrokering (see messaging service). This does not use Web Services as such at present but the filters can be controlled by HPSearch services.
Messaging Service	This is used to stream data in workflow fed by real-time sources. It is based on NaradaBrokering which can also be used in cases just involving archival data
Notification Service	This supplies alerts to users when filters (data-mining) detects features of interest
QuakeTables Database Services	The USC QuakeTables fault database project includes a web service that allows you to search for Earthquake faults.
Scientific Plotting Services	We are developing Dislin-based scientific plotting services as a variation of our Web Map Service: for a given input service, we can generate a raster image (like a contour plot) which can be integrated with other scientific and GIS map plot images.
Data Tables Web Service	We are developing a Web Service based on the National Virtual Observatory's VOTables XML format for tabular data. We see this as a useful general format for ASCII data produced by various application codes in SERVO and other projects.
Key interfaces/standards/software Used	GML WFS WMS WSDL XML Schema with pull parser XPP SOAP with Axis 1.x UDDI WS-Context JSR-168 JDBC Servlets WS-Management VOTables in Research
Key interfaces/standards/software NOT Used (often just for historical reasons as project predated standard)	WS-Security JSDL WSRF BPEL OGSA-DAI

Key GIS and Related Services

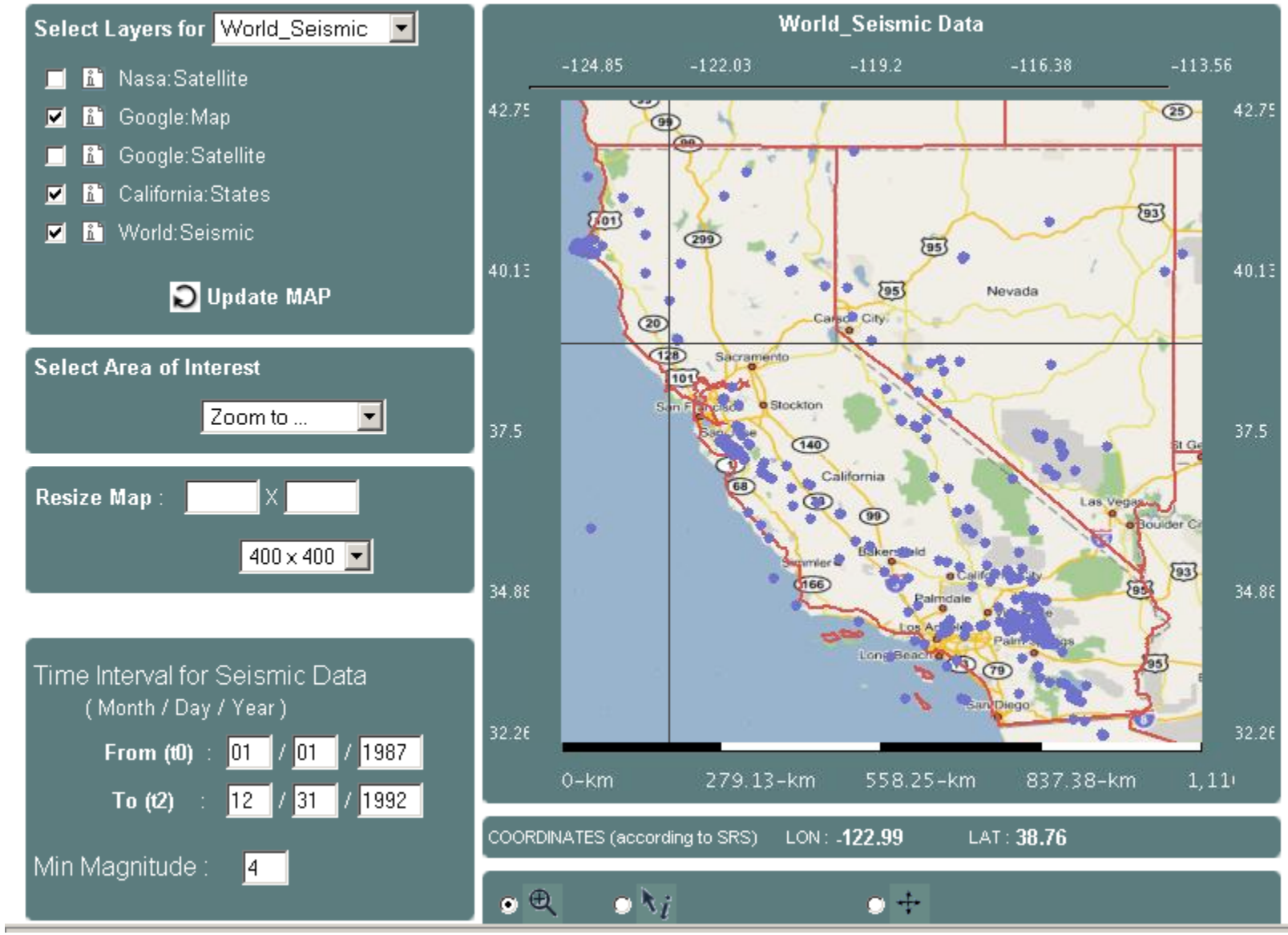
Component	Description
HPSearch	Support for streaming data between services; supports scriptable workflows so not limited to DAGs; implementation of WS-Distributed Management
WS-Context	Contexts can be used to hold arbitrary content (XML, URIs, name-value pairs); can be used to support distributed session state as well as persistent data; currently researching scalability.
Web Feature Service	Supports both streaming and non-streaming returns of query results.
Web Map Services	Supports integration of local and remote map services; treats Google maps as an OGC-compliant map server;
Sensor Grid	Publish/subscribe system allows data streams to be reorganized using topics.

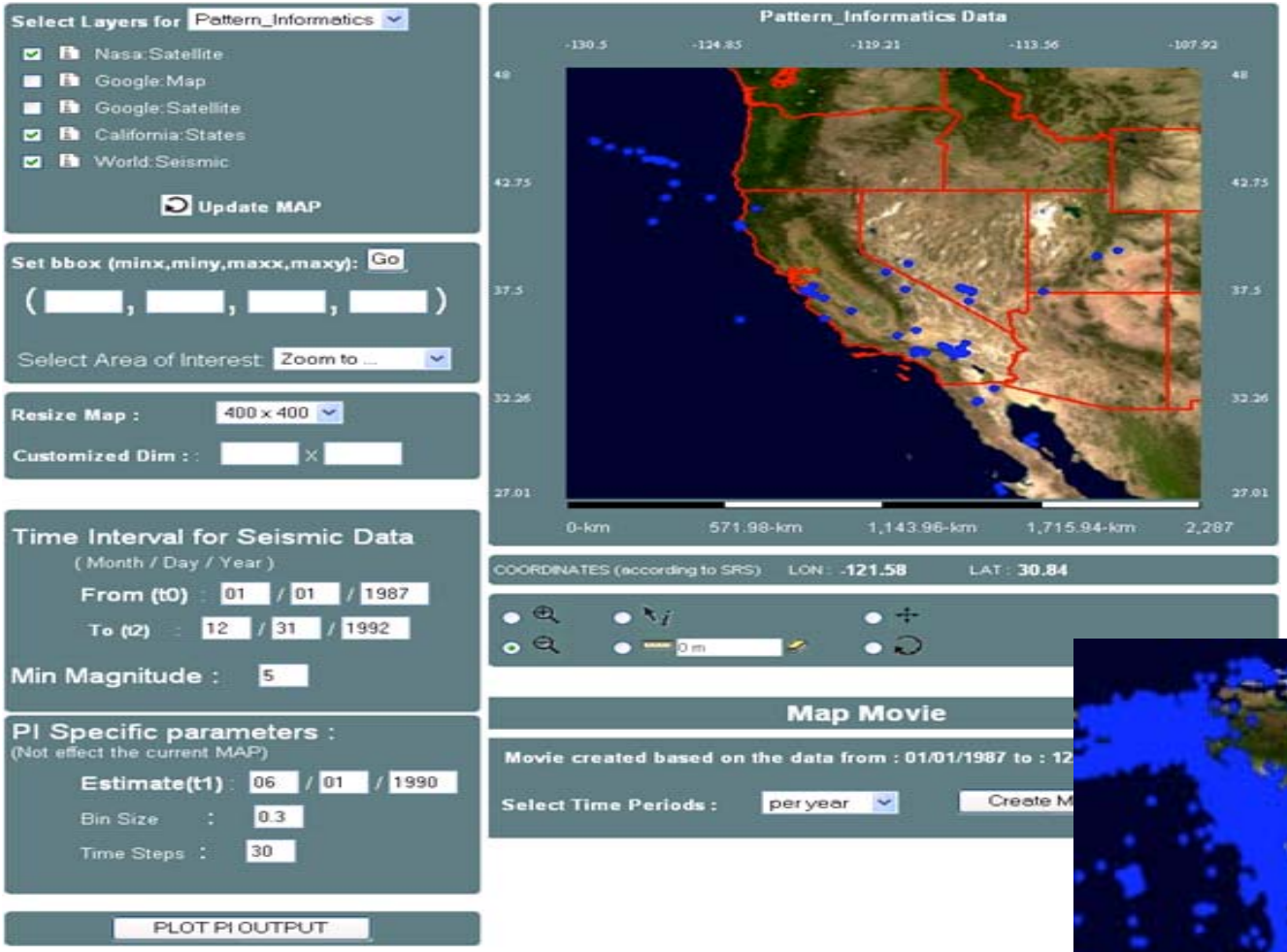
GISand Geophysical Applications

- Pattern Informatics
  - Earthquake forecasting code developed by Prof. John Rundle (UC Davis) and collaborators.
  - Uses seismic archives as input
- Regularized Dynamic Annealing Hidden Markov Method (RDAHMM)
  - Time series analysis code by Dr. Robert Granat (JPL).
  - Can be applied to GPS and seismic archives.
  - Can be applied to real-time data.
- Virtual California
  - Prof. Rundle's UC-Davis group
  - Used for simulating time evolution of fault systems using fault and fault friction models.
  - Assimilation version uses seismic archives

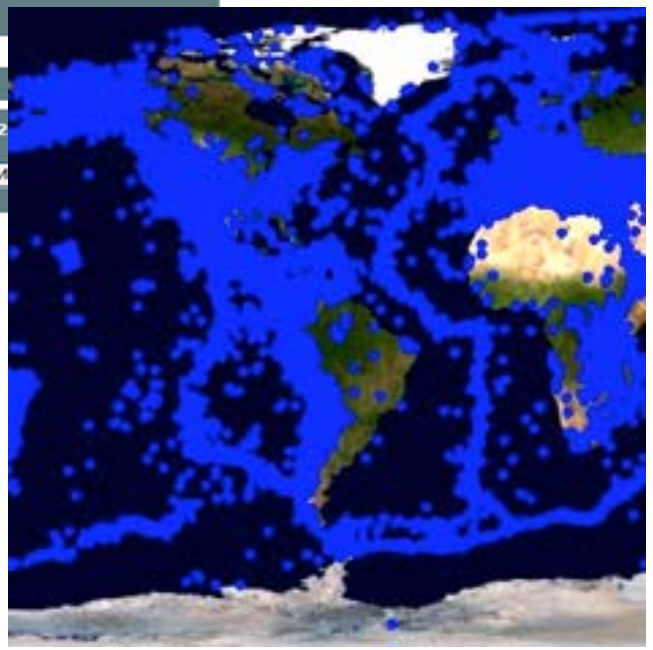
Pattern Informatics

- This has been our simplest “proving ground” example.
- Integrates (streaming) WFS, WMS, WS-Context, and HPSearch’s WSProxy services (wraps PI executable and helper format conversion services).
- This is basically a linear workflow

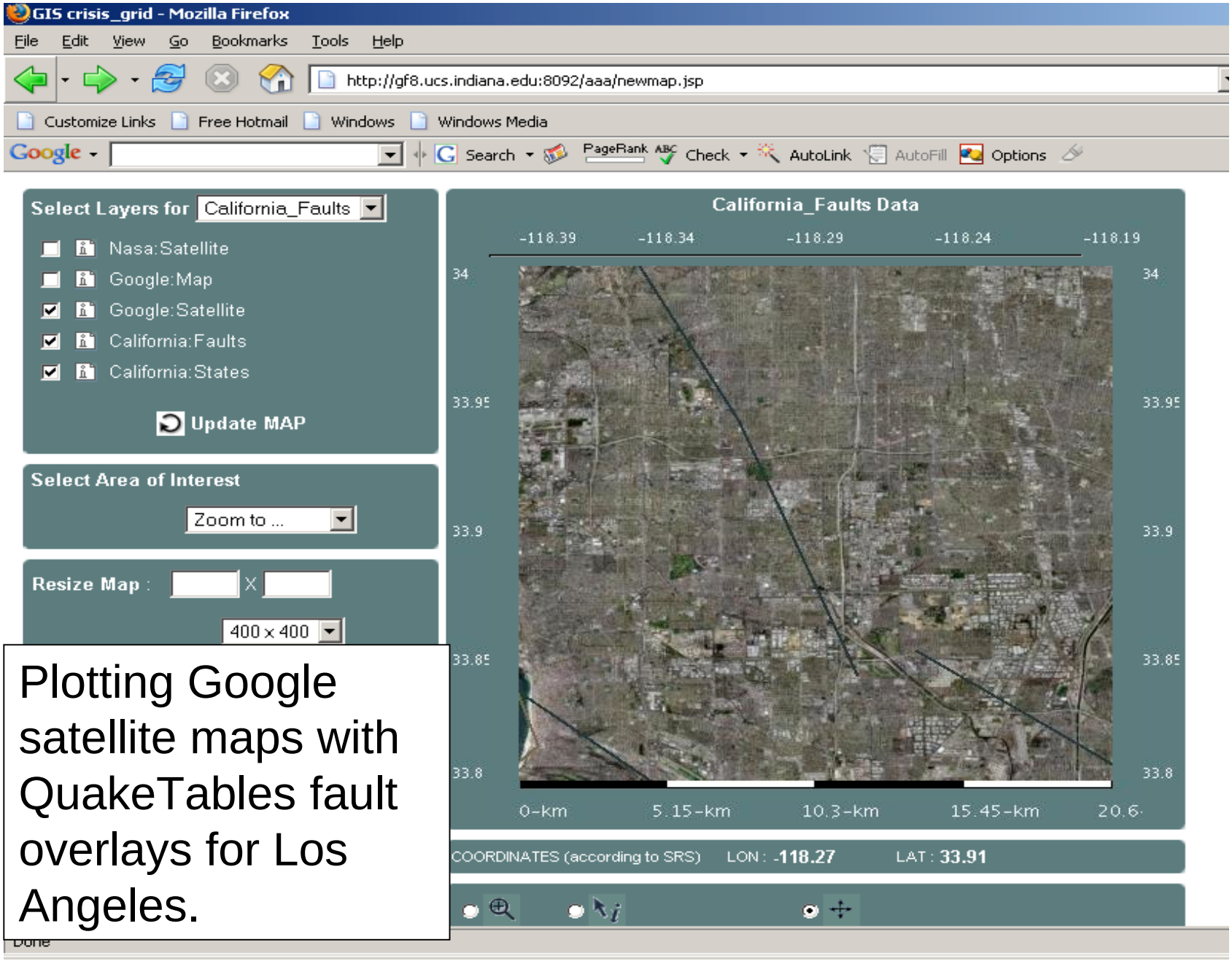

Select Layers for World_Seismic
Nasa:Satellite
Google:Map
Google:Satellite
California:States
World:Seismic
Update MAP
Select Area of Interest
Zoom to ...
Resize Map :
X
400 x 400
Time Interval for Seismic Data
(Month / Day / Year)
From (t0) : 01 / 01 / 1987
To (t2) : 12 / 31 / 1992
Min Magnitude : 4
World_Seismic Data
-124.85
-122.03
-119.2
-116.38
-113.56
42.75
40.13
37.5
34.86
32.26
Nevada
California
Sacramento
Stockton
Las Vegas
Palmdale
San Diego
0-km
279.13-km
558.25-km
837.38-km
COORDINATES (according to SRS) LON : -122.99 LAT : 38.76



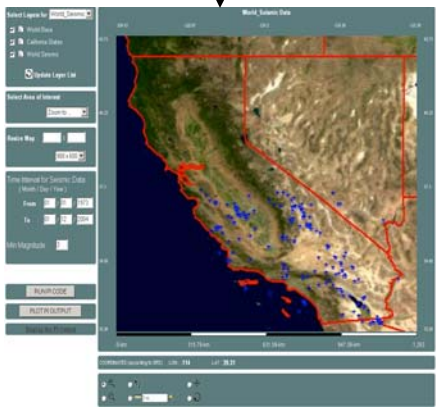
Pattern informatics results combined with Feature and Map servers can be used to forecast areas of increased earthquake probability.



Whole earth seismic catalog plotted on NASA map server. Combines streaming feature server and map server.



Plotting Google satellite maps with QuakeTables fault overlays for Los Angeles.



Web Map Client

Stubs

WSDL

Aggregating WMS

Stubs

SOAP

HTTP

WSDL

WFS + Seismic Rec.

WSDL

WFS + State Bounds

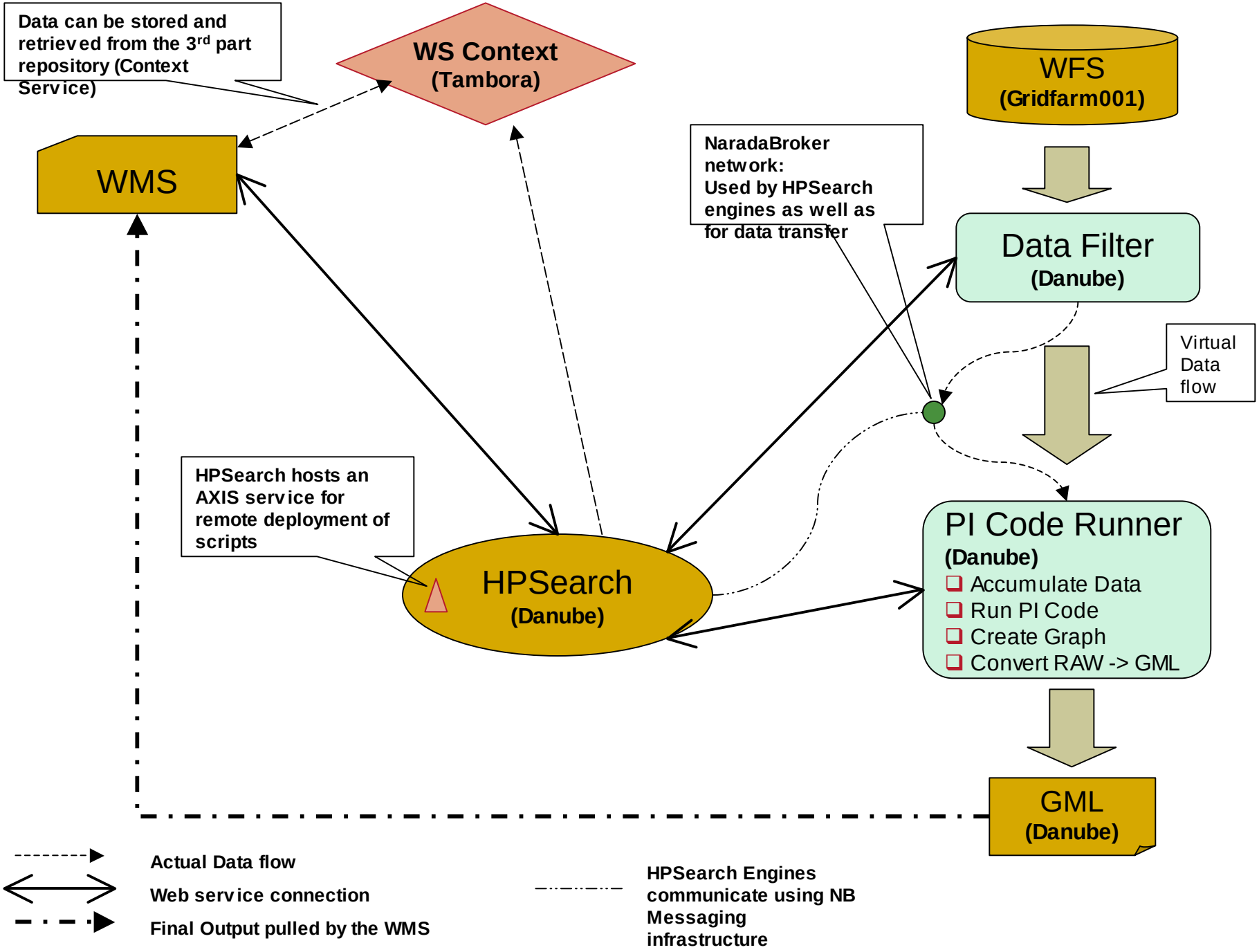
. . .

“REST”

WMS + OnEarth Or Google Maps

Tying It All Together: HPSearch

- HPSearch is an engine for orchestrating distributed Web Service interactions
  - It uses an event system and supports both file transfers and data streams.
  - Legacy name
- HPSearch flows can be scripted with JavaScript
  - HPSearch engine binds the flow to a particular set of remote services and executes the script.
- HPSearch engines are Web Services, can be distributed interoperate for load balancing.
  - Boss/Worker model
- ProxyWebService: a wrapper class that adds notification and streaming support to a Web Service.
- More info: http://www.hpsearch.org

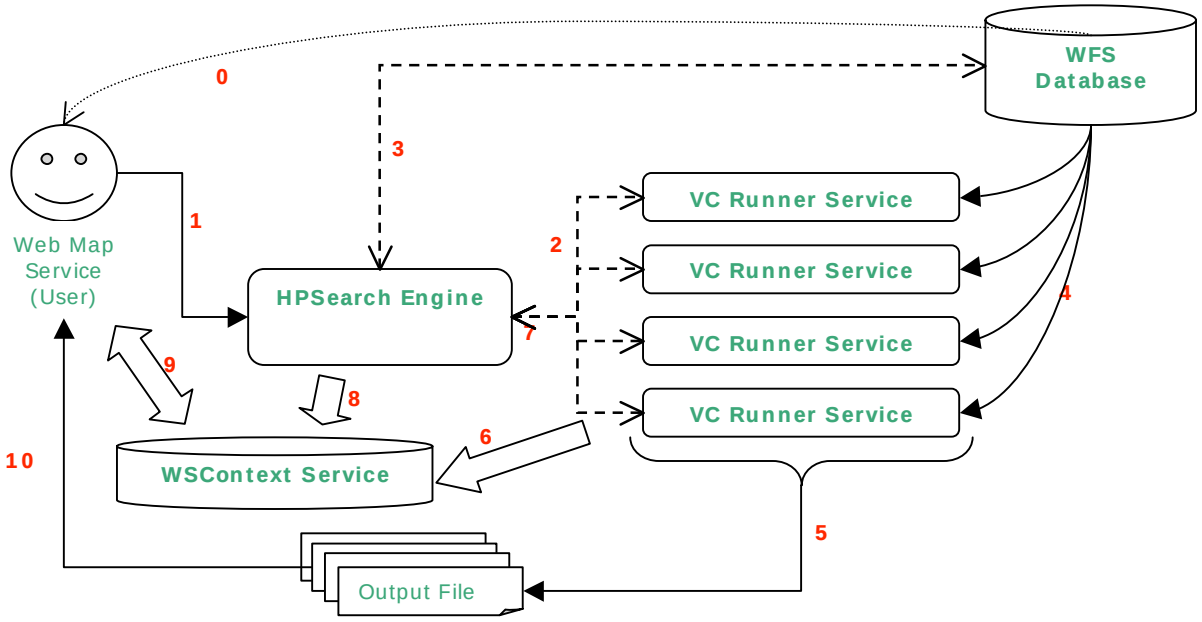

Data can be stored and retrieved from the 3rd part repository (Context Service)
WS Context
(Tambora)
WFS
(Gridfarm001)
WMS
NaradaBroker network:
Used by HPSearch engines as well as for data transfer
Data Filter
(Danube)
Virtual Data flow
HPSearch hosts an AXIS service for remote deployment of scripts
HPSearch
(Danube)
PI Code Runner
(Danube)
Accumulate Data
Run PI Code
Create Graph
Convert RAW -> GML
GML
(Danube)
Actual Data flow
Web service connection
Final Output pulled by the WMS
HPSearch Engines communicate using NB Messaging infrastructure

Scripting Workflows

- HPSearch allows web service interactions to be scripted.
- The script for the PI interaction is available from www.hpsearch.org/demos/PI/PICodeRunner.js.html
- Other scripts are available from http://www.hpsearch.org/demos/index.html

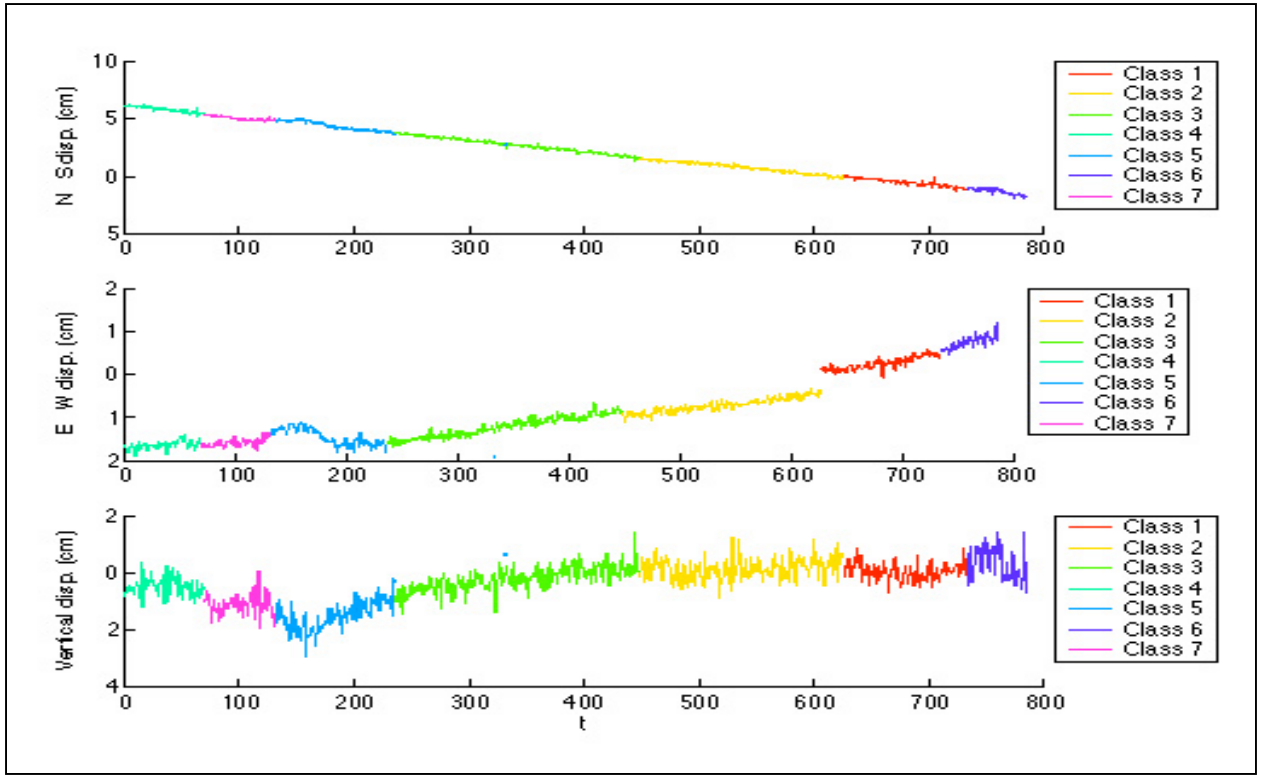
Virtual California

- The data assimilation version of VC uses uncoupled “worker” nodes to model the California seismic record.
- The best worker survives and is used for a forecast.
- This shows a simple parallel workflow with HPSearch.
- WMS, WFS, WS-Context all used as before.



RDAHMM: GPS Time Series Segmentation

Slide Courtesy of Robert Granat, JPL



GPS displacement (3D) length two years.

Divided automatically by HMM into 7 classes.

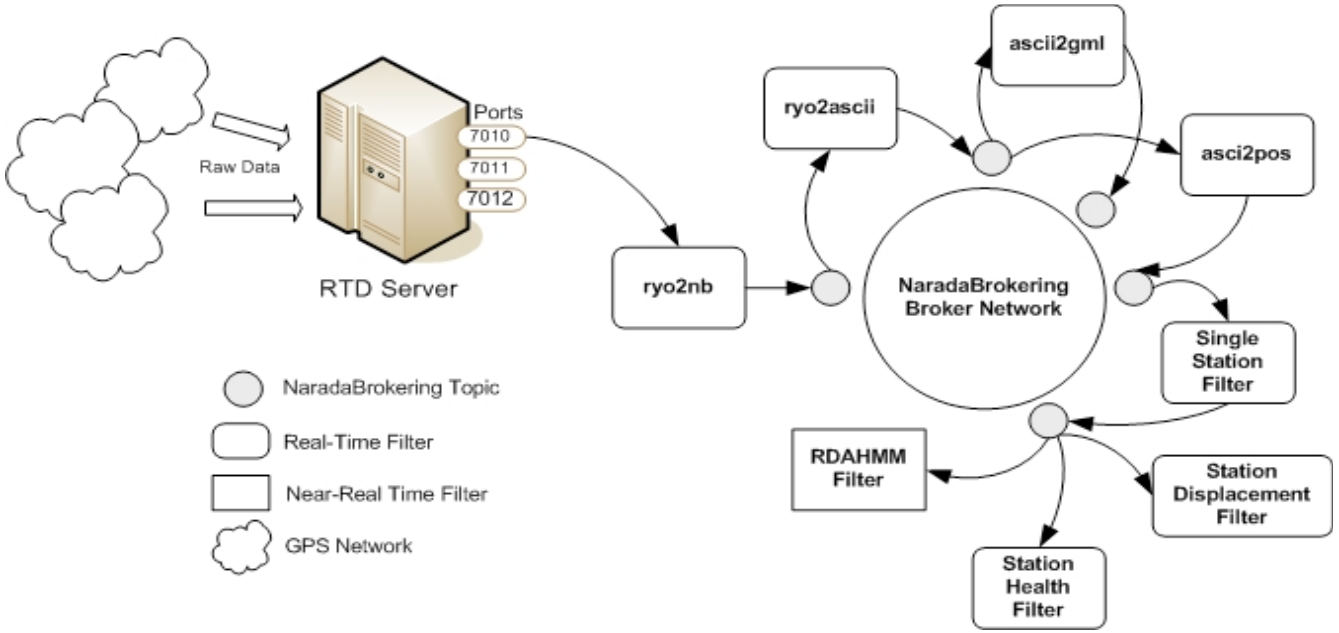
Features:

- Dip due to aquifer drainage (days 120-250)
- Hector Mine earthquake (day 626)
- Noisy period at end of time series

- Complex data with subtle signals is difficult for humans to analyze, leading to gaps in analysis
- HMM segmentation provides an automatic way to focus attention on the most interesting parts of the time

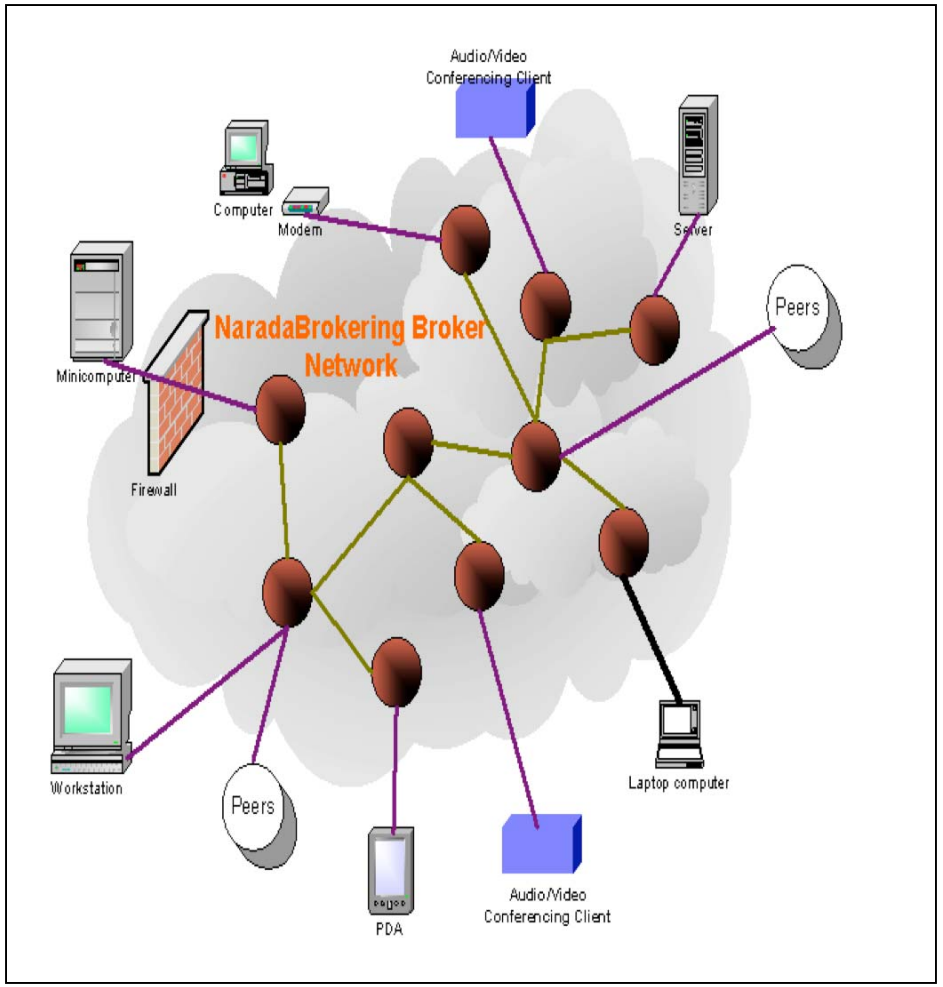
RDAHMM and Real Time GPS

- Illustrates integration of sensor grid services with
- Streaming GPS position data is successively filtered until we separate out the individual station's stream
- Topic-based publish/subscribe is used to manage streams and filters.
- Filters are Web services; use HPSearch for stream management.
- RDAHMM is treated as another filter.



NaradaBrokering: Message Transport for Distributed Services

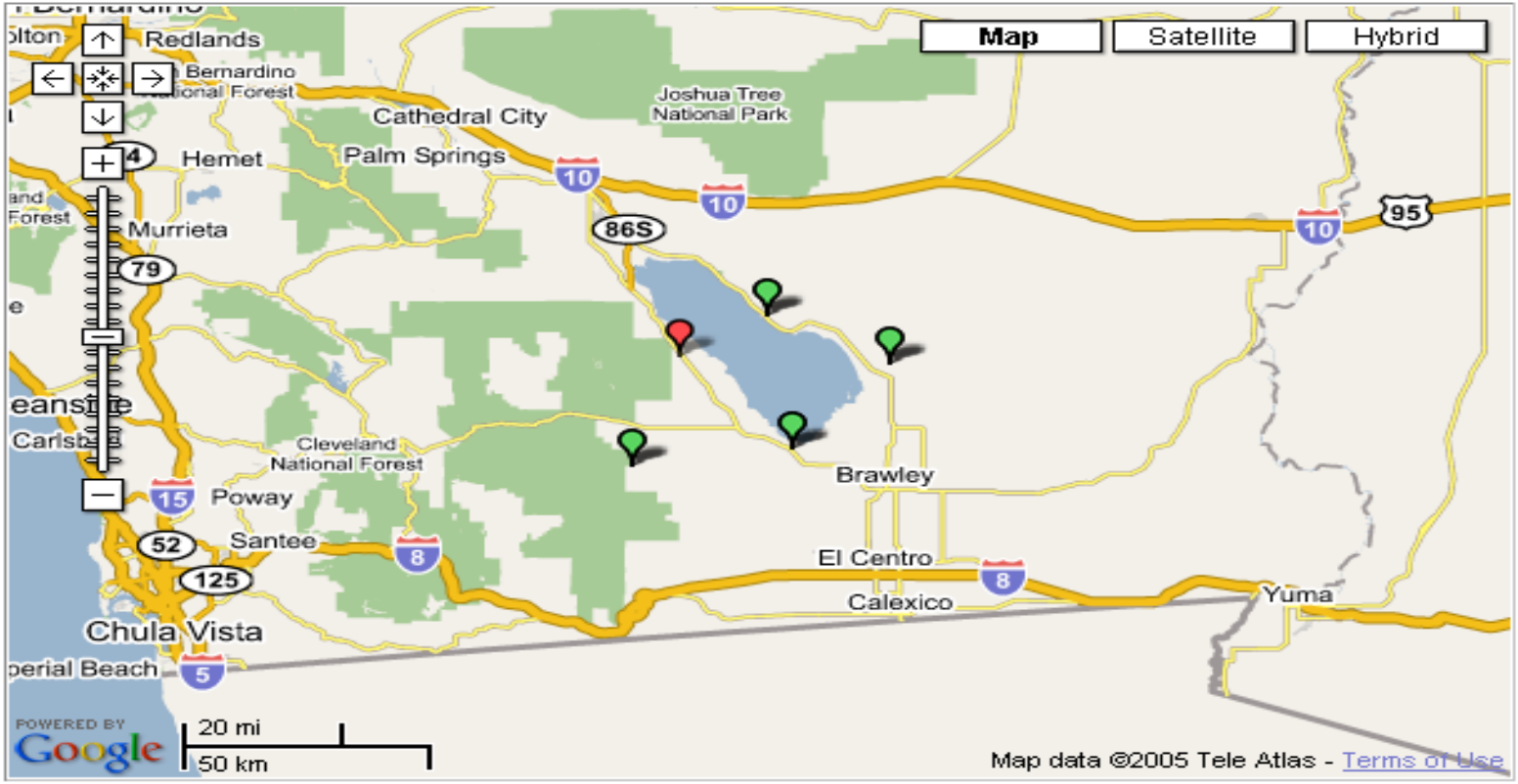
- NB is a distributed messaging software system.
  - http://www.naradabrokering.org
- NB system virtualizes transport links between components.
  - Supports TCP/IP, parallel TCP/IP, UDP, SSL, GridFTP
- See e.g. http://grids.ucs.indiana.edu/ptliupages/publications/AllHands2005NB-Paper.pdf for trans-Atlantic parallel tcp/ip timings.



Customize Links | Free Hotmail | Windows | Windows Media

Google | Search | PageRank | Check | AutoLink | AutoFill

IMPERIAL GPS Network
Real-Time Position Messages

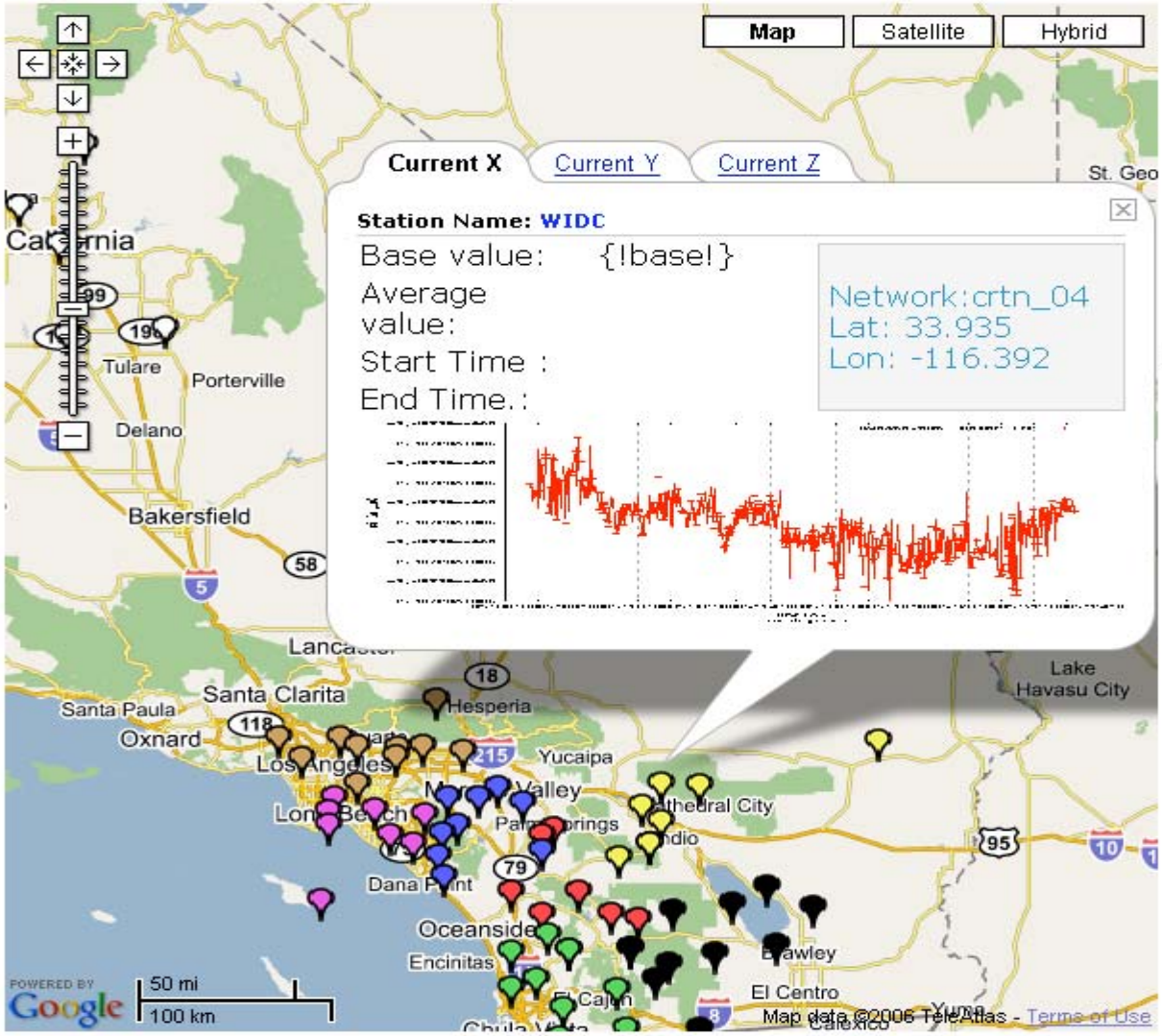


Station Name	Latitude	Longitude
CRRS	33.06980331667044	-115.73502754597638
GLRS	33.27480829562368	-115.52135766047812
USGC	33.030058944755766	-116.08532015921372
DHLG	33.389801746972516	-115.78801461129845

Done

SOPAC Real Time GPS Networks

Click on a station symbol for more information.



Network Icon Color

- crtn_03
- crtn_01
- crtn_05
- crtn_02
- cvsrn
- crtn_04
- crtn_06
- crtn_07

Real-time GPS position data created by filters is available through Google Map interface.

Sample RDAHMM reports should be available by end of Summer.

More information about California Real Time Network (CRTN) is available at SOPAC Web Page

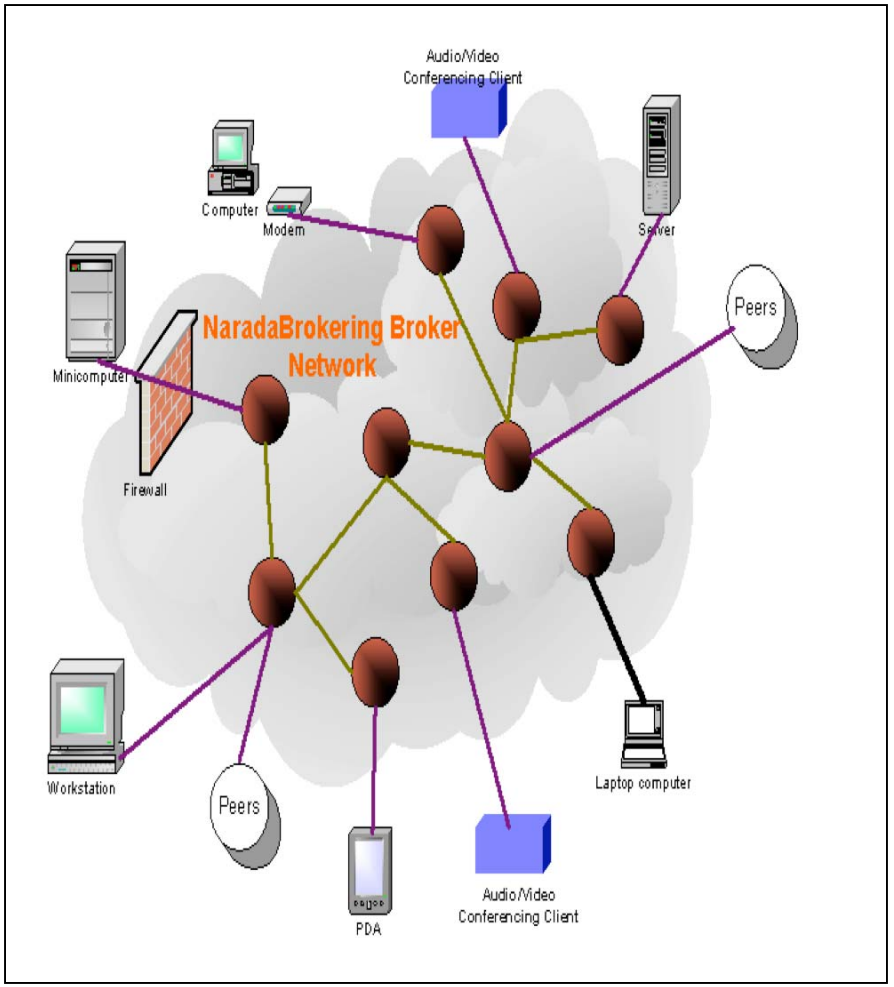
More Information

- Contact me: mpierce@cs.indiana.edu
- Portal:
  - http:complexity.ucs.indiana.edu:8282
- QuakeSim web site
  - http://quakesim.jpl.nasa.gov/
- Project Wiki: www.crisisgrid.org
- Registered SourceForge project
  - http://sourceforge.net/projects/crisisgrid

Additional Slides

NaradaBrokering: Message Transport for Distributed Services

- NB is a distributed messaging software system.
  - http://www.naradabrokering.org
- NB system virtualizes transport links between components.
  - Supports TCP/IP, parallel TCP/IP, UDP, SSL, GridFTP
- See e.g. http://grids.ucs.indiana.edu/ptliupages/publications/AllHands2005NB-Paper.pdf for trans-Atlantic parallel tcp/ip timings.



More Info on GPS Stations

Network Name	RTD Server Address	Stations
LACRTN	132.239.154.69:5014	VTIS, HBCO, CVHS, LORS, TABL, UCSB, AZU1, CSDH, DYHS, VDCY, UCLP, CIT1, LAPC
PARKFIELD	n/a	HOGS, POMM, MIDA, CRBT, CARH, LAND, MNMC, LOWS, RNCH, CAND, MASW, TBLP, HUNT
OCRTN	132.239.154.69:5010	OEOC, CAT2, WHYT, TRAK, SACY, MJPK, SCMS, SBCC, FVPK, BLSA
SDCRTN	132.239.154.69:5013	P486, MONP, RAAP, MVFD, P472, SIO5, DVLW, PMOB, P480, DSME, OGHS
IMPERIAL	132.239.154.69:5012	SLMS, CRRS, USGC, DHLG, GLRS
DVLRTN	132.239.152.72:8001	DVLE, DVNE, DVSW, DVSE, ESRW, DVLS, DVNW, ESE2
CVSRN	132.239.154.69:5015	COMA, RBRU, LEMA
RCRTN	132.239.154.69:5011	PIN2, WIDC, KYVW, PSAP, COTD, PIN1, MLFP, CNPP, BILL, EWPP, AZRY

Example Topic Fields

Network Name	RYO Topic (null filter Publishes to)	ASCII topic (ryo2ascii filter Publishes to)
LACRTN	/SOPAC/GPS/LACRTN/RYO	/SOPAC/GPS/LACRTN/ASCII
PARKFIELD	/SOPAC/GPS/PARKFIELD/RYO	/SOPAC/GPS/PARKFIELD/ASCII
OCRTN	/SOPAC/GPS/OCRTN/RYO	/SOPAC/GPS/OCRTN/ASCII
SDCRTN	/SOPAC/GPS/SDCRTN/RYO	/SOPAC/GPS/SDCRTN/ASCII
IMPERIAL	/SOPAC/GPS/IMPERIAL/RYO	/SOPAC/GPS/IMPERIAL/ASCII
DVLRTN	/SOPAC/GPS/DVLRTN/RYO	/SOPAC/GPS/DVLRTN/ASCII
CVSRN	/SOPAC/GPS/CVSRN/RYO	/SOPAC/GPS/CVSRN/ASCII
RCRTN	/SOPAC/GPS/RCRTN/RYO	/SOPAC/GPS/RCRTN/ASCII

SERVO Tools in Other Projects

- WS-Context, Web Feature Service, and Web Map Service used as part of Homeland Security demo at Los Alamos National Lab D division.
  - Integrated with IEISS simulation code for modeling electrical and natural gas grids.
- WS-Context and HPSearch are being used in the NSF ITR funded VLAB project.
- We are working with JPL and Scripps/UCSD to deploy our sensor grid services with RDAHMM and S_T Filter.
- Currently investigating Web Map Service integration and performance caching for Indiana.